PB85-183002

# Legal and Administrative Impediments to the Conduct of Epidemiologic Research

Technical Resources, Inc., Bethesda, MD

Prepared for

Environmental Protecton Agency, Washington, DC

Dec 84

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-183002

# Legal and Administrative Impediments to the Conduct of Epidemiologic Research

December 1984

Sponsored by the

**TASK FORCE ON**

**ENVIRONMENTAL CANCER AND HEART AND LUNG DISEASE**

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

50272-101

| REPORT DOCUMENTATION PAGE | 1. REPORT NO. | 2. | 3. Recipient's Accession No. PB85 183002/AS |
|---|---|---|---|
| 4. Title and Subtitle: Legal and Administrative Impediments to the Conduct of Epidemiologic Research | | | 5. Report Date: December 1984<br>6. |
| 7. Author(s): Task Force on Environmental Cancer and Heart and Lung Disease | | | 8. Performing Organization Rept. No. |
| 9. Performing Organization Name and Address: Technical Resources, Inc.<br>3202 Monroe Street, Suite 300<br>Rockville, MD 20852 | | | 10. Project/Task/Work Unit No.<br>11. Contract(C) or Grant(G) No.<br>(C) 68-02-3775<br>(G) |
| 12. Sponsoring Organization Name and Address: U.S. Environmental Protection Agency<br>Task Force on Environmental Cancer and Heart and Lung Disease<br>401 M Street, SW.; Office of Research & Development (RD-672)<br>Washington, DC 20460 | | | 13. Type of Report & Period Covered: Final Report<br>14. |

15. Supplementary Notes

Task Force Chairman, William Ruckelshaus, Administrator
Task Force Working Group Chairman, Scott Baker, Ph.D.

16. Abstract (Limit: 200 words)

This report was sponsored in response to concerns expressed in the Fourth, Fifth, and Sixth Annual Reports of the Task Force on Environmental Cancer and Heart and Lung Disease. The Annual Reports recommended that legal impediments to the conduct of both Federal and non-Federal epidemiologic research be studied. To determine whether significant modifications or relaxation of impediments to epidemiologic research have occurred since 1981, an assessment of the current situation was undertaken. The study methodology consisted of a search and review of recent literature and interviews of individuals from Federal organizations involved with these issues. The findings indicate that (1) legislative changes have not occurred to significantly modify existing privacy and confidentiality laws or restrictions on use of Agency data, (2) several Federal Agencies are exploring ways to increase access to Federal data and allow record linkages of specifically defined data elements for approved epidemiologic research projects, (3) resource-use data are not available to quantify the added costs of current research impediments, and (4) even if legal and administrative impediments are relaxed, information processing barriers and limitations to the most efficient use and linkage of existing Federal data may continue to hamper the conduct of epidemiologic research studies.

17. Document Analysis a. Descriptors

epidemiology, medical records, mortality, vital statistics

b. Identifiers/Open-Ended Terms

scientific research (legal aspects), record linkage

c. COSATI Field/Group: Biological and Medical Sciences/Clinical Medicine

| 18. Availability Statement: Release unlimited | 19. Security Class (This Report): Unclassified | 21. No. of Pages: 53 |
|---|---|---|
| | 20. Security Class (This Page): Unclassified | 22. Price |

(See ANSI-Z39.18) See Instructions on Reverse

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce

# Legal and Administrative Impediments to the Conduct of Epidemiologic Research

December 1984

Sponsored by

## TASK FORCE ON

## ENVIRONMENTAL CANCER AND HEART AND LUNG DISEASE

U.S. Environmental Protection Agency

Public Health Service

National Cancer Institute • National Heart, Lung, and Blood Institute

National Institute for Occupational Safety and Health

National Institute of Environmental Health Sciences

National Center for Health Statistics

Centers for Disease Control • Food and Drug Administration

Department of Energy • Consumer Product Safety Commission

Occupational Safety and Health Administration • Department of Agriculture

Copies of this document may be obtained free-of-charge through requests to

Technical Resources, Inc.
3202 Monroe Street, Suite 300
Bethesda, Maryland 20852
(301-231-5250)
Attention: Mary Jane Arnold

When the supply is exhausted, the document may be purchased from

National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22161

CONTENTS


| | Page |
|---|---|
| PREFACE | ii |
| EXECUTIVE SUMMARY | iii |
| LEGAL AND ADMINISTRATIVE IMPEDIMENTS TO THE CONDUCT OF EPIDEMIOLOGIC RESEARCH | 1 |
| Importance of Epidemiologic Research to Current Public Health Status | 4 |
| Legal and Administrative Impediments to Efficient and Cost-Effective Conduct of Epidemiologic Research Projects | 6 |
| Impact of Legal and Administrative Impediments | 11 |
| Existing Information Collected on Individuals by Federal Agencies | 17 |
| Importance of Record Linkage in Conducting Epidemiologic Research | 20 |
| Implementation Difficulties | 23 |
| Current Activities in Linking and Improving Access to Federal Information | 26 |
| Quantification of the Costs of Legal Impediments | 28 |
| Conclusion | 30 |
| PROPOSALS TO IMPROVE THE CONDUCT OF EPIDEMIOLOGIC RESEARCH | 32 |
| REFERENCES CITED | 37 |
| BIBLIOGRAPHY | 40 |
| Appendix A LIST OF INTERVIEWEES | |
| Appendix B GLOSSARY | |

# PREFACE

This study of legal and administrative impediments to the conduct of epidemiologic research was performed under the auspices of the Task Force on Environmental Cancer and Heart and Lung Disease, mandated by Congress in Public Law 95-95, Section 402. The Task Force Working Group requested this study to determine if significant modifications or relaxations of impediments to data access and record linkage have occurred since the Task Force recommendation on this issue in its 1981 Fourth Annual Report to Congress.

The study was conducted by Dr. Maria Pavlova of the U.S. Environmental Protection Agency and Mr. Paul Young. The report was reviewed by all interviewees and relevant comments integrated and incorporated. However, the views and conclusions do not necessarily reflect opinions of individual interviewees and should not be construed as representing the official position of any Agency or organization.

The Task Force expresses its appreciation to all the interviewees (listed in Appendix A) for their cooperation, time, and sincere interest in this study. A special thanks to Dr. Gilbert Beebe of the National Cancer Institute for permission to use his previously published studies and for his time and effort in multiple interviews and report reviews. The Task Force is especially indebted to Dr. James Whittenberger, Director of the Southern Occupational Health Center at the University of California at Irvine, for his expert assistance in the final review, editing, and revision of this report. Finally, the authors thank Ms. Mary Jane Arnold and Ms. Joanna Fringer of Technical Resources, Inc., for their meticulous editing and patient guidance of the report through numerous review cycles.

## EXECUTIVE SUMMARY

Epidemiologic research plays a crucial role in the study of environmentally related diseases. When multiple valid epidemiologic studies show a positive association between a specific environmental exposure and a human disease, their findings tend to dominate decisions concerning the environmental agent's toxicity and the need for regulations. Hence, the importance of epidemiologic findings dictates that epidemiologic research be conducted as efficiently and effectively as possible. For more than a decade, formal and informal discussions have occurred concerning access to data about individuals by authorized scientific researchers conducting approved studies in the interest of improving public health.

This report, discussing the current state of legal and administrative impediments to the conduct of epidemiologic research, addresses the following major issues:

1. Changes that may have occurred between 1980 and 1983 to legal impediments that hamper the access to data on individuals for epidemiologic research studies

2. Current Federal administrative and legislative activities directed to permitting greater access to existing Federal data bases and permitting record linkages of data

3. The need for and feasibility of quantifying the additional costs incurred in conducting epidemiologic research studies because existing legal and administrative barriers hamper the effective use of Federal data

4. Identification of the practical administrative barriers and access limitations that would continue to exist even if legal and administrative impediments were modified and relaxed.

The study methodology consisted of a search and review of recent literature on these issues; interviews of Federal epidemiologists, identified by Task Force members, who are actively conducting epidemiologic research or are involved with these issues; and interviews of persons from the Office of Health Planning and Evaluation, Public Health Service (PHS), the Office of Public Affairs, the Department of Health and Human Services (DHHS), the General Accounting Office (GAO), the Office of Management and Budget (OMB), the Social Security Administration (SSA), and the Internal Revenue Service (IRS).

The findings of the study indicate that

- Legislative changes have not occurred to modify existing privacy and confidentiality laws or Agency data use restrictions.
- Several Federal administrative activities are in progress to permit increased access to Federal data for Institutional Review Board-approved epidemiologic research projects and to allow record linkages of specifically defined data elements that can identify a particular individual.
- Resource-use data are not available to quantify the added costs of current research impediments; thus, if Congress and the public are to be convinced that legislative action is needed to improve epidemiologic research effectiveness, a cost study may be required.
- Even if legal and administrative impediments are relaxed, information-processing barriers and limitations to the most efficient use and linkage of existing Federal data may continue to hamper the conduct of epidemiologic research studies.

With appropriately controlled access through selected integrated record linkages, a health researcher could monitor specifically predefined

data from existing Federal data bases for variations that suggest opportunities for reducing health hazards, reflect the influence of prevention and control programs, and point to additional opportunities to learn more about the etiology of major diseases.

To achieve these ends, the following legislative actions are proposed:

- Provide access to Federal tax records maintained by the SSA and IRS for authorized Federal health researchers to obtain necessary linkage data for epidemiologic research cohort and case control subject identification
- Extend access to the IRS Individual Master File for last known filing address and date of filing to DHHS and other Federal health Agencies authorized to conduct epidemiologic research
- Modify the Privacy Act (as recommended by the 1977 President's Privacy Protection Study Commission) to permit interagency record data linkages under proper safeguards, to improve the effectiveness and efficiency of conducting epidemiologic research
- Support the National Death Index, implemented and maintained by the National Center for Health Statistics, to ensure its long-term continuation and improvement.

These proposals are discussed in detail, with their supporting rationales, following the body of the report.

# LEGAL AND ADMINISTRATIVE IMPEDIMENTS TO THE CONDUCT OF EPIDEMIOLOGIC RESEARCH

For more than a decade, epidemiologists and other medical researchers have attempted to inform state and national lawmakers about the importance of relaxing privacy and confidentiality legislation in one specific area: namely, to provide access to data about individuals to authorized scientific researchers conducting approved medical and health investigations in the interest of improving public health. Some of the more recent formal attempts are as follows:

- Report of the Privacy Protection Study Commission (1977)
- Report of the Committee on Data Protection (1978)
- Report of the Interagency Task Force on Ionizing Radiation (DHEW 1979)
- Report by the Subcommittee on Oversight and Investigations of the Committee on Interstate and Foreign Commerce, House of Representatives, 96th Congress "Data Transfer Restrictions Impede Epidemiological Research" (December 1980)
- Annual Reports to Congress of the Task Force on Environmental Cancer and Heart and Lung Disease (1981, 1982, 1983)
- Environmental Health--A Study of the Issues in Locating, Assessing, and Treating Individuals Exposed to Hazardous Substances (DHHS/NCHS 1981)
- Banbury Report No. 9: Quantification of Occupational Cancer (Cold Spring Harbor Laboratory 1981)
- Cancer Risk--Assessing and Reducing the Dangers in Our Society (Office of Technology Assessment 1982).

These reports, as well as many articles in scientific journals, point out the importance of epidemiologic research to improvements in prevention and control of disease and the negative impact of restricted access to information on the conduct of epidemiologic research. For example, the Tax Reform Act of 1976 (P.L. 94-455) severely restricted access by health researchers to the existing IRS Individual Master File for economical tracing of the most current residence of individuals or next-of-kin in cohort and case-control studies. Tracing of individuals to determine health status and to identify possible disease causal factors is a critical component of epidemiologic research. Tracing is presently very time-consuming and often expensive. Prior to this Act, epidemiologists in health Agencies had broad access to the latest filing address of individual taxpayers through the IRS files at a cost of about $0.10 per name.

The Tax Reform Act was modified by Public Law 96-128 to facilitate research involving military veterans. It allows Federal Agencies, other than the National Institute for Occupational Safety and Health (NIOSH), to obtain address information from IRS files to conduct Institutional Review Board (IRB)-approved health research projects dealing with military personnel who may have been exposed to hazardous substances during active service. Also, cooperative research projects with NIOSH can utilize this limited IRS file access authorization. (NOTE: NIOSH researchers can obtain a taxpayer's address; but not the last filing date from which the address was obtained).

This report, compiled from interviews and literature reviews, describes the current state of legal and administrative impediments to epidemiologic research. The study revealed

- In the past several years, there have been no significant reductions in legal and administrative impediments to access and linkage of information regularly collected by Federal Agencies on individuals.
- The only recent major occurrence that has affected the efficiency of conducting epidemiologic studies is implementation of the National Death Index (NDI) data base by the National Center for Health Statistics (NCHS). The implementation of this important national data base will reduce both the time and cost of identifying a case-control or cohort subject's vital status, especially as the data base expands in future years. (The NDI contains deaths that occurred during or after 1979.)

Because legal impediments have not changed significantly, this report reviews the following:

1. Importance of epidemiologic research in identifying and monitoring disease prevention and control measures
2. Legal and administrative impediments to efficient and cost-effective conduct of approved epidemiologic research projects
3. Existing information collected on individuals by Federal Agencies
4. Importance of record linkages in conducting epidemiologic research
5. Current activities in linking Federal information and permitting improved access to such information
6. Quantification of the costs of legal and administrative impediments.

## Importance of Epidemiologic Research to Current Public Health Status

Improvements in public health services (sanitation, immunizations, health education) and medical care have greatly reduced and, in some instances, eliminated infectious diseases that previously killed large numbers of people in infancy and during childhood. Now many more people live to old age, when cancer, heart, and lung diseases cause significant mortality (OTA 1982, 4). The complex interaction of exposures to environmental pollution occurring over a lifetime and lifestyle factors associated with the development of chronic and irreversible changes in metabolic functions and body organs and tissues must be addressed by the use of all the tools of medical research, including laboratory experimentation, clinical study, and epidemiologic investigation. The research findings will necessarily make important contributions to prevention and control strategies and will help to set environmental and occupational exposure standards.

Epidemiology is the study of the distribution and determinants of disease in populations. It is an observational science useful for identifying and assessing the adverse results of certain life experiences, thereby providing clues to disease etiology, for example, the role of environmental factors in the onset of cancer (Newill 1983, 215). In addition to identifying disease risks, epidemiology may play a positive prevention role by identifying aspects of diet or behavior that are protective against disease. In an era of increasing awareness of the interplay between environment and health, the need for information on environmental factors influencing health far outstrips its availability (Beebe 1983, 76).

Epidemiologic findings play a crucial role in the regulatory process. For example, when epidemiologic findings from multiple valid studies show a positive association between an exposure or behavior and cancer in numans, these findings tend to dominate decisionmaking about carcinogenicity (OTA 1982, 11). Regulators feel more comfortable, as does Congress, if the effects of environmental agents have been demonstrated in human populations (Radford 1983, 205). The Congressional reliance on epidemiologic research findings is amply demonstrated by the recent introduction of several bills and the passage of several public laws directing conduct of epidemiologic studies:

- Public Law 95-601, to do epidemiologic research on the health effects of low-level ionizing radiation
- Senate Bill S.1865, 96th Congress (1980), introduced but not passed, to require the Public Health Service (PHS) to conduct a comprehensive assessment of adverse health effects from the Atomic Weapons Test Program
- House Bill HR.6745, 96th Congress (1980), introduced but not passed, to establish the most likely statistical relationship between a given exposure and the consequent harmful effects therefrom, on the basis of epidemiologic data and other scientific information
- Public Law 95-623, to do epidemiologic research to identify the environmental conditions that are responsible for causing or aggravating human diseases, and specifically to determine whether cancer, birth defects, genetic damage, emphysema, asthma, bronchitis, heart disease, stroke, and mental illness represent such diseases or effects.

While, on the one hand, Congressional representatives are urging and even insisting that epidemiologists tackle all of these important problems, on the other hand, Congress has enacted legislation that restricts access to the information that may be required. For some types

of epidemiologic studies, the epidemiologist enters the scientific and associated political and environmental policy arena with one hand tied behind his or her back.

## Legal and Administrative Impediments to Efficient and Cost-Effective Conduct of Epidemiologic Research Projects

The field of epidemiologic research comprises several types and subtypes of study methodologies, all of which are adversely affected, to varying degrees, by legal and administrative barriers to information collection. Before discussing the forms of impediments, it is important to understand the different types of study methodologies.

For the purpose of this discussion, epidemiologic studies are divided into three general types: (1) experimental, (2) descriptive (ecologic), and (3) observational (analytic). Observational epidemiology is, in turn, subdivided into three subtypes: (a) cohort studies, also sometimes called longitudinal or prospective studies; (b) case-control studies, also referred to as retrospective or case-referent studies; and (c) cross-sectional studies, sometimes called prevalence studies. While several basic strategies exist, there are no rigid study designs within any of these categories.

### Experimental Epidemiology

Experimental epidemiology, which involves the deliberate application or withholding of a factor followed by observation for an effect, is an ideal procedure for investigating cause-and-effect hypotheses. However, ethical considerations preclude purposefully exposing people to suspected harmful environmental agents, especially given the severity and irreversibility of

cancer, heart disease, and lung disease. Experimental epidemiology is being constantly practiced, on the other hand, to test measures or interventions deemed possibly or probably beneficial or thought to aid in disease prevention (OTA 1982b, 137). These type studies are more commonly known as "clinical trials."

Intervention studies are difficult to conduct because of the need to secure the cooperation of a large group of people willing to permit an experimenter to intervene in their lives. Experimental epidemiologic studies have been performed, but with the current emphasis on prevention programs, particularly those involving lifestyle factors (diet, nutrition, and exercise), it is likely that even more studies will be conducted or sponsored in the form and context of clinical trials. The National Heart, Lung, and Blood Institute (NHLBI) has already conducted several large clinical trials, including the Multiple Risk Factor Intervention Trials (MRFIT), the Hypertension Detection and Follow-Up Program (HDFP), and lipid studies.

Descriptive Epidemiology

Descriptive epidemiology examines the distribution and extent of disease in populations according to basic characteristics, e.g., age, sex, and race. The primary purpose of conducting descriptive epidemiologic studies is to obtain clues about the etiology of a disease, which may then be investigated more thoroughly in analytic studies. Examination of age-specific rates of disease occurrence or mortality across time are examples of descriptive epidemiology (OTA 1982b, 137). Perhaps the best known descriptive epidemiologic studies are the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) national program (NCI 1981)

and the NCI Epidemiology Branch county mortality pattern studies with resulting Cancer Atlases (Mason, McKay, Hoover et al. 1975, 1976).

Because descriptive epidemiology deals with the statistics of incidence and mortality on a class rather than individual basis, it is only marginally affected by legal impediments to data acquisition. Both state and Federal Agencies are much more willing to share statistically aggregated data, minus individual identifiers, for further aggregation into regional or national class sets.

The primary legal and administrative impediments to epidemiologic research relate to using the general findings from a descriptive epidemiologic study to then conduct a more detailed study of the individuals constituting the specific class at high risk. This latter-type study falls into the third class, observational epidemiology.

Observational Epidemiology

Observational epidemiology depends on data derived from observations of individuals to determine if an association exists between a suspected factor and a disease and, if so, the strength of the association based on accepted statistical methods (OTA 1982b, 137). The types of observational epidemiologic studies, cohort, case-control, and cross-sectional, differ in the population groups selected for study.

A cohort study starts with a group of people, called a cohort, considered free of the disease under study (at the time of selection),[1] who are defined by some common characteristic (e.g., born during the

1. This is often a historical cohort (i.e., one established on the basis of historical records) whose health status is then ascertained and monitored in the present timeframe (Robert S. Gordon, Jr., interview, National Institutes of Health, July 5, 1983).

same specified time period, or living in a common locale, or exposed to a suspect adverse environmental factor, or evidencing similar personal attributes or behavior, such as diet). The group is then studied over time and the health status of the individual members observed. Thus, in the cohort study, the researcher has developed a hypothesis of specified potential causes and is looking for effects (Lloyd 1983, 12).

The impact of legal and administrative impediments on the conduct of cohort studies is substantial and extends to various aspects:

1. Obtaining identifying information on persons having a common attribute (e.g., born near the Love Canal site in Niagara Falls, New York)

2. Identifying other variables and confounding factors that must be evaluated to develop conclusions (work history by labor classification and industry)

3. Obtaining access to record systems to trace individuals to determine current disease incidence or mortality.

Although much of this information presently is collected by the Federal Government, it resides in a variety of different Agencies and legally cannot be easily linked.

In a case-control study, individuals with the disease or condition under study (cases) are assembled and their past or present attributes are compared with an equal or greater number of individuals without the disease or condition (controls) to identify relevant risk factors (Sartwell 1983, 27). Case-control studies (that are not mortality-based studies) are much less affected by legal and administrative impediments than cohort studies. Cases and controls are usually obtained from hospital medical record files, and while administrative and perceived legal barriers

may exist, no current Federal laws restrict researcher access to this information.[2] An example of a perceived legal barrier is the tightening of restrictions on access to medical records imposed by hospitals. This tightening was precipitated by the privacy laws, which responsible hospital personnel erroneously interpreted as applying to all data about an individual (Beebe 1983a, 10).

If individuals identified at the outset of the study are alive, legal impediments to locating their current residences usually are not a problem. This is particularly true for cancer and heart and lung diseases, which are chronic in nature and usually require regular, periodic follow-up by a physician. Thus, primary information on risk factors can be obtained from the subjects themselves, with a subject's consent from employers and hospitals, or second-hand from immediate relatives and friends.

Quality assurance of data is a vital concern and the inaccuracies inherent in second-hand information are well-documented. Therefore, when researchers must use second-hand information, they also search for additional sources of information, particularly risk factor data. Many of these additional or secondary information sources are Federal Agency data files (e.g., industry of employment from SSA records, military occupational exposures from Department of Defense (DoD) records, lifestyle factors from NCHS

---

2. If a particular case-control study depends on historical records for determining exposure or occupation and this information is only available from records such as those maintained by the SSA, the impact of legal and administrative impediments on the conduct of the study is similar to that encountered in conducting a cohort study. The same may apply to case-control studies conducted within a population identified in a cohort study. These are known as "nested" case-control studies (David P. Brown, National Institute for Occupational Safety and Health, telephone interview, Cincinnati, Ohio, September 13, 1983).

health surveys) that could confirm important facts, efficiently and cost effectively, if it were not for current legal and administrative restrictions on access.

The third type of observational study is cross-sectional. Cross-sectional studies provide information on prevalence of disease in a population at one specific point in time (Higgins 1983, 37). Exposures, personal characteristics, and biological effects are all measured at the same time. Each individual is examined only once and there is no provision for follow-up. Cross-sectional studies may be aimed simply at fact-finding or occasionally used to test a hypothesis (Sartwell and Last 1980, 46). Such studies have provided useful information on environmental and occupational hazards among limited groups of people, such as the discovery of a large number of cases of pneumoconiosis in tin smelters (Higgins 1983, 40).

For cross-sectional epidemiologic studies, in which a sample of the general population or a particular community or a special subgroup (e.g., schoolchildren, workers in a particular industry) is selected, without regard to risk factors or disease incidence and with no programmed follow-up, the impact of legal and administrative impediments to the conduct of the study is minimal. Supplementary information may be sought through records of medical care or occupational exposure, and because the subject is voluntarily participating in the survey, permission for access can be readily obtained.

## Impact of Legal and Administrative Impediments

As is evident from the above discussion, most types of epidemiologic research projects can be conducted successfully and effectively without

major concern for current legal and administrative impediments. It is only the cohort observational studies and certain forms of historical case-control studies[3] that are severely impaired by legal and administrative impediments on access to and linkage of specific data collected by Federal Agencies and identifiable to a particular individual. Unfortunately, cohort epidemiologic studies may be the most valuable and perhaps the only effective means for establishing

- Systematic monitoring and surveillance systems for persons exposed to an increasing variety and complexity of toxic substances, both on the job and in their everyday lives
- Feedback systems to chart the progress of prevention and control measures over long time periods (Beebe 1983a, 4).

The advantages of cohort studies are that (1) they allow observation of all outcomes, not only those originally anticipated; (2) they can yield incidence rates as well as relative risk; and (3) study bias in ascertaining risk factors is reduced because classification into an exposure category cannot be influenced by prior knowledge that the disease exists. The disadvantages are that cohort studies require large numbers of subjects; a long follow-up period; and comprehensive information data bases linking work history, episodes of exposure, medical history and contacts with different parts of the health care system, lifestyle factors and behaviors, and ultimately cause of death information (OTA 1982b, 138-139). As will be discussed in the next section, much of this information already

---

3. The conduct of case-control studies would be greatly enhanced in efficiency and effectiveness if retrospective work histories for individuals were readily available from sources such as the SSA.

is collected by Federal Agencies for large segments of the population. The problem, then, is aggregating the necessary specific information into a single system for identifiable individuals. In the absence of a system for such linkages, clarification of causal factors in some environmentally related chronic diseases may not be possible. Because the technology exists to perform the linkage, the legal and administrative impediments are the crux of the issue.

Probably the most definitive papers to date on the need for and barriers to linkage of various work, health, and death records are those published by Dr. Gilbert W. Beebe, "Record Linkage and Needed Improvements in Existing Data Resources" from the 1981 Banbury Report No. 9: Quantification of Occupational Cancer and "Long-Term Follow-Up is a Problem" from the American Journal of Public Health, March 1983, Vol. 73, No. 3. The following excerpts from these papers and statements from recent interviews illustrate the impact of the major legal and administrative impediments:

> The Privacy Act of 1974 (P.L. 93-579) and the later Tax Reform Act of 1976 (P.L. 94-455) have had a chilling effect on medical research both directly, in their restriction of access to Federal record systems, and indirectly, through their ripple effects in state legislatures and private institutions (particularly hospitals and nursing homes) (Beebe 1983b, 245-246).
>
> The sharing of individual identifying records among Agencies is severely restricted by legal statutes, especially those pertaining to the Census Bureau (13 U.S.C. 8,9), the Internal Revenue Service (26 U.S.C. 6103), the SSA (42 U.S.C. 1306), and the NCHS (42 U.S.C. 242m(d)) (Beebe 1981, 666).
>
> Two major sources of mortality information--the Social Security Administration files and the Veterans Administration (VA) files-- were greatly impaired by the 1981 tax bill that curtailed eligibility for the lump-sum death benefit programs of SSA, VA, and other Agencies beginning October 1, 1981. This tax action eliminated the need for the death reports that previously had been required to be filed to obtain the benefit. Had the National

Death Index not been put in place by NCHS before the curtailment took place, there would have been no systematically collected source of mortality follow-up after 1981 (Beebe 1983b, 245-246).

Under the Paperwork Reduction Act of 1978 (P.L. 96-511) and in the interests of simplifying Government forms, the occupation of taxpayer and spouse requested on every type of Form 1040 was removed from the new Form 1040EZ, first used in 1983, thereby eliminating a potentially valuable piece of linkage data for future studies (Frederick Scheuren, Internal Revenue Service, telephone interview, Washington, D.C., May 13, 1983).

Because of the Federal Freedom of Information Act (FOI) versus state privacy of information statutes dilemma, Federal health researchers have difficulty obtaining death certificates from individual states to classify cause of death for research subjects. The dilemma is that Federal researchers cannot fully protect death certificates from disclosure because of FOI regulations. However, unless Federal researchers can guarantee they will never give out the certificates, the states will not provide the death certificates, because of state concerns for violation of their own privacy regulations. NCHS, however, can provide protection from FOI disclosure under the law cited above (42 U.S.C. 242m(d)) (Aaron Blair, National Cancer Institute, interview, Bethesda, Maryland, April 15, 1983).

SSA collects annual earnings reports from all employers that contain employee Social Security Number (SSN) and other selected information of value in conducting occupational epidemiologic studies. These SSA records are declared to be IRS tax records and therefore usually are not available to health researchers for development of industry cohorts. However, if written authorization is given by the individual company executives, the data can be released to researchers (Beebe 1981, 667).

The Freedom of Information Act has worked in reverse for Federal Agency epidemiologic researchers because of the perception that the Act requires the Government to always provide collected information in response to public inquiries from legal and medical personnel, thereby making

- States reluctant to provide death certificates or workmen's compensation claim data files to qualified researchers
- Institutions reluctant to provide medical record and disease diagnosis data
- Industries reluctant to provide union membership, exposure, or disability claim information (Aaron Blair, National Cancer Institute, interview, Bethesda, Maryland, April 15, 1983).

It is important to note that this perception is incorrect and researchers can protect study data under several directives and Office of the General Counsel (OGC) rulings (John P. Fanning, Department of Health and Human Services, interview, Washington, D.C., September 8, 1983).

As is evident from the above list, the epidemiologist attempting to conduct a long-term cohort study of a large number of persons to identify risk factors associated with the potential onset of cancer, heart and lung disease, and other environmentally related diseases may have to go to great expense in time and money to get information that has already been collected.

Cost Impact

The quantifiable cost impact of these legal and administrative impediments is evident from an illustration presented in a keynote speech by Dr. Robert S. Gordon, Jr., Special Assistant to the Director, National Institutes of Health. In 1981, at the opening of the Task Force Workshop on Exposure to Environmental Agents, Their Metabolism, and Mechanisms of Toxicity, he stated the following:

> An investigator now has to pay Equifax or another such Agency $100 or so per name to have a cohort traced, instead of the $0.10 per name that it formerly cost to run the computer files at IRS to get the most recent mailing addresses for a list of people. That is a real difference, a matter of a thousand-fold in the cost of locating people whose current address you may need for very valuable research (Task Force 1981, A14).[4]

---

4. IRS confidentiality provisions have been modified to provide certain qualified investigators with access to taxpayers' addresses, but only (1) to NIOSH for the purpose of locating individuals who are known to, or may have been specifically exposed to occupational hazards or (2) through NIOSH for veterans who are being specifically studied for occupationally related exposure and disease risk, in order to determine the status of their health (P.L. 95-210). A recent interview with senior IRS personnel has verified that the actual cost for obtaining these addresses is still about $0.10 per name.

Social Impact

The social impacts are potentially much greater. Although the results of epidemiologic studies are now commonly reported in the press and on television, it is probable that the general public and even Congress do not understand that because the health consequences of environmental exposures often are long-delayed and not uniquely associated with a single agent, record linkage is often the only way of establishing relationships between exposure and disease (Beebe 1981, 666).

As many as 880,000 American workers, or slightly more than one percent of the current labor force of 84 million persons, presently experience full- or part-time exposures to carcinogens regulated by the Occupational Safety and Health Administration (OSHA) (NIOSH 1977, 3). Furthermore, OSHA regulates only a small number of the more than 55,000 chemicals in current use in commerce. While there is controversy surrounding the estimates of workplace-associated cancer, most estimates range from 5 to 15 percent of the 400,000 cancer deaths per year (OTA 1982a, 88). This is 20,000 to 60,000 cancer deaths alone, without considering possible mortality and morbidity from heart, lung, and other diseases (e.g., neurotoxic or reproductive effects) induced by occupational exposure.

Many persons suggest that in epidemiologic research prime consideration be given to a national disease incidence surveillance network for occupational groups, with particular emphasis on occupations where potentially hazardous exposures are known to be high and continuous (DHEW 1977, 425). While an effective surveillance system covering this segment of the population could be implemented, it would require a substantial

number of informational items from a variety of existing national data bases, addition of currently uncollected information for the occupational work force, and a sophisticated mechanism for controlled integrated record linkage.

Recommendations for strengthening existing systems, for developing new data sources, for creating data set linkages, and for resolving data access problems are contained in a November 1983 Report of the Occupational Cancer Risk Subcommittee on Occupational Cancer Estimation (DHHS/CCERP 1983). Such an integrated data system would not only serve a major practical need in occupational health but also provide (1) a warning mechanism for less prominent, but related health effects in the general population and (2) a much needed mechanism to reflect the influence of rapidly expanding prevention and control programs (DHEW 1977, 425). Thus, substantial numbers of deaths and illnesses might be prevented with proper control techniques, if the causal relationships could be more efficiently determined by record linkage.

## Existing Information Collected on Individuals by Federal Agencies

The literature sources cited in the bibliography provide an extensive list and description of information bases of potential value for selected data item record linkage. These include

- National Center for Health Statistics
  - Health and demographic surveys
  - Health manpower inventories and surveys
  - National Vital Statistics System
  - Health resources utilization statistics
  - National Death Index

- National Cancer Institute

  - SEER data base of cancer incidence, survival, and mortality on approximately 10 percent of the population

- National Institute for Occupational Safety and Health, Centers for Disease Control

  - NIOSH National Occupational Hazard Survey (NOHS)
    Note: These survey data are presently being updated in a new effort called the National Occupational Exposure Survey (NOES)

- Social Security Administration

  - One percent Continuous Work History Sample (CWHS)
  - Industry annual earnings reports by employee (from employers)
  - Disability Insurance Fund
  - Numident and Alphadent Files created from SSN applications (Form SS-5)

- Bureau of the Census, Department of Commerce

  - 1980 Census of Population "long form" used for about 20 percent of the population
  - Current Population Survey (CPS)
  - March Annual Supplement to the Post Enumeration Survey (PES) for the 1980 census

- Internal Revenue Service

  - Form 1040 series of Individual Taxpayer Returns (for only a few specifically identified nonsensitive items such as occupation)
  - Forms W-2, W-2p and 1099 from employers
  - Statistics of Income (SOI) sample

- Bureau of Labor Statistics, Department of Labor

  - Health and safety statistics
  - Day-to-day logs of occupational illness and injury

- U.S. Department of Agriculture

  - Pesticide surveys
  - Nutrition surveys

- Health Care Financing Administration
  - Medicare Application File
  - Medicare Payments File
  - Medicaid Summary File
- Veterans Administration and Department of Defense
  - Retired military and disabled veterans health care records
  - Retired military and disabled veterans mortality records
  - Mortality records for 95 percent of all wartime active-duty persons.

As is evident from the above list of record titles, even without a description of specific data elements, the Federal Government collects a vast array of epidemiologically important personal data: birth date and place, occupation, industry of employment, general demographics, health care, morbidity, and mortality. In addition, much of this collection has been going on for decades, so that there already exists a wealth of information that only needs to be made available and linked to help provide invaluable information concerning the etiology of some of our major diseases. As Dr. Beebe stated,

> The long-term prospects [for national data on occupational mortality] are at least good,...because the great need for the information will eventually be recognized by Congress and the upper levels of Government, and most of all because all the data elements of a system already exist--they need only be joined together in a system (Beebe 1983, 76).

## Importance of Record Linkage in Conducting Epidemiologic Research

An often quoted definition of record linkage is as follows:

> Each person in the world creates a book of life. This book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this book into a volume (Sartwell and Last 1980, 31).

Technically, a record linkage system makes it possible to relate significant health events in an individual's life that are remote from one another in time and place. The prerequisites are a procedure for uniquely identifying individuals even if they change their names, a method of abstracting and storing relevant information about them from health and vital records, and a technique for error-free matching of bits of information obtained from different sites and settings over prolonged periods. The final step is output of statistical tables. If all of this can be efficiently and successfully accomplished, the result is one of the most powerful tools available for epidemiologic studies. Record linkage systems with these qualities have been operational for some years in the Oxford area in England, Scotland, Sweden, and, on a smaller scale in some other parts of the world (Sartwell and Last 1980, 31).

The most important aspect of record linkage that must be understood and accepted by policymakers and Federal and state lawmakers is that

> without individual identifying information, which is essential for follow-up and selected record linkages, it is virtually impossible to carry out epidemiologic studies of the etiologic characteristics, risk factors, natural history, and prognosis of disease; nor can studies be conducted in which new approaches to prevention, early detection, treatment, and delivery of health services are developed (Gordis and Gold 1980, 154).

The linkage and availability of identifying data on a specific individual obviously raise serious concerns about privacy and confidentiality and the potential for abuse. However, epidemiologists are health professionals who recognize and adhere to ethical standards in research to the same high degree as other doctors, nurses, and allied health professionals. While there is no formal confidentiality oath required, a confidentiality pledge is certainly implied every time a research protocol is approved by an Institutional Review Board (IRB).

To ensure uniformity of confidentiality procedures throughout the Federal Government, the Science Advisor to the President, Dr. George A. Keyworth, is presently obtaining final Agency concurrence on a model policy on the Protection of Human Research Subjects prior to Federal Register publication in early 1985. This model policy would require all Federal Departments or Agencies to adopt as a common core the regulations governing research with human subjects issued by the DHHS (codified as 45 CFR, Part 46), as periodically amended or revised, while permitting additions needed by any Department or Agency that are not inconsistent with the core provisions. These core regulations prescribe specific steps and procedures that must be followed by the principal investigator and approved by the IRB for conduct of epidemiologic research involving linkage of data that can be identified to a specific individual. Such procedures require that the individual be informed of the extent to which confidentiality of records identifying the subject will be maintained (OPRR 1983, 17-28). As a practical matter, epidemiologists do assure their subjects of the degree of confidentiality they legally can offer.

These three characteristics of epidemiologic research--(1) the professional ethics of the researchers themselves, (2) the requirement to incorporate confidentiality provisions and procedures into the research protocol before it will receive IRB approval, and (3) the ability of the human research subject to expressly state the degree of confidentiality he or she wants maintained, including denying collected data to be used for record linkage--enable the noted epidemiologists Sir Richard Doll and Richard Peto to state:

> We know of no instance in any country in the world where the provision of information for epidemiological research has led to any abuse of privacy and many instances where the lack of such information has led to unnecessary disease (Doll and Peto 1981, 1239).

U.S. Government records contain a wealth of information about individuals that, if specifically defined and linked, could be of great value to researchers looking for associations between exposures and disease incidence and mortality. For example, to conduct a study relating mortality to occupational exposures, "occupation" could be extracted from the IRS record and linked to "employer" and "industry" information collected by the SSA.[5] "Cause of death" can already be obtained from the death certificate filed in the state in which the death is recorded, using the NCHS National Death Index to locate quickly and inexpensively the appropriate state (OTA 1982b, 151). Unedited information on primary occupation and industry of employment can also be obtained from the death

---

5. The IRS has completed a feasibility study of this procedure. By linking this information, IRS has been able to code occupation and create a dictionary of occupational titles coded to different levels of SOC codes.

certificate and would be a secondary means of data verification. Thus, an inexpensive preliminary historical cohort study could be conducted using specifically identified elements from existing data files to survey whether a particular occupation within a particular industry has a statistically higher risk of disease, without going to the time and expense of contacting the industry, unions, and other data sources to build the more traditional cohort. If the preliminary survey appears to show a positive correlation for the disease, SSA files might be used in conjunction with IRS files to identify in a cost-effective manner a suitable cohort (complete with current last filing addresses) for a more comprehensive follow-up investigation of specific risk factors.

## Implementation Difficulties

While the need for record linkages should be evident from the above discussion and the linkages that would be most advantageous appear obvious, the interviews and literature reviews disclosed several technical barriers and practical limitations to access and linkage of information over extended timeframes:

- The IRS does not routinely code the occupation of the taxpayer given on the Form 1040 series; in fact, as stated above, the new Form 1040EZ does not even contain space for recording occupation.

- The IRS keeps only three years of coded data on readily accessible files, after which these data are kept for another three years on magnetic tape, and then transferred to microfilm, which is sent to the various regional service centers for final storage. Thus occupation would have to be matched on a current basis to SSA records to be readily available in future years for historical cohort studies. Because the IRS does not currently enter occupation from the tax forms on the Individual Master File, there is an additional cost to implement such a

procedure (Marjory Karlsen, Internal Revenue Service, telephone interview, Washington, D.C., May 9, 1983).

- SSA individual work history data must be obtained by conducting a lengthy manual search of boxed (pre-1950) and microfiched (post-1950) original records. The current costs for this process are about $2.00 per subject per quarter of traced employment. Thus, it currently is not practical to obtain work history data for a large cohort (Warren Buckler, Social Security Administration, telephone interview, Baltimore, Maryland, June 30, 1983).

- Decennial Census data lack Social Security Numbers; therefore, accurate matching tends to be limited to contemporary records for which a date of birth and current residence address can be used (Beebe 1981, 668).

- While standard death certificates have a space for the SSN, a usable SSN is recorded on only about 80 percent of certificates (Beebe 1981, 669).

- Although the existing National Death Index's matching algorithms are fairly effective in generating positives, the user is responsible for assessing the quality of each match and weeding out false positives. Except for very large studies, a substantial portion of the false positives can be eliminated easily by careful inspection of the NDI output. This, however, is potentially time-consuming for users submitting large volumes of records and requires such users to develop their own computerized techniques for eliminating false positives. NCHS is presently studying the existing matching algorithms to minimize the generation of false positives (Letter from Manning Feinleib, National Center for Health Statistics, report review comments, Hyattsville, Maryland, July 8, 1983).

- SSA collects information on industry for all those in covered employment but occupation only for those who file for disability awards.

- While all but one of the 10 SEER registries collect information on occupation from the hospital record, this has not been required for the NCI aggregated national data (Beebe 1981, 669).

- The NCHS Hospital Discharge Survey does not include information on occupation and industry (Beebe 1981, 669).

- Although NCHS health and demographic surveys do not routinely collect SSN for persons interviewed, NCHS is currently collecting SSN in the Hispanic National Health and Nutrition Examination Survey and will collect SSN in future National Health Interview Surveys (Manning Feinleib, National Center for Health Statistics, interview, Hyattsville, Maryland, April 8, 1983).

- The National Vital Statistics System, which is a cooperative statistical system between NCHS and the states, provides critical data on births and deaths that are used in epidemiologic studies (Manning Feinleib, National Center for Health Statistics, interview, Hyattsville, Maryland, April 8, 1983). The quality and timeliness of the data, especially cause of death, are in constant jeopardy due to inadequate Federal funding.

- The NCHS automated system (ACME) that produces statistics on the underlying cause of death is also a unique epidemiologic resource (Manning Feinleib, National Center for Health Statistics, interview, Hyattsville, Maryland, April 8, 1983). While the ACME system is currently used in 15 states, its implementation in additional states is constrained by the lack of available resources.

- NCHS, in cooperation with the Bureau of the Census and NIOSH, has demonstrated the feasibility of coding occupation and industry from the death certificate using standardized procedures. These procedures have now been adopted by 15 states. Expansion to other states and incorporation of this activity into the national vital statistics cooperative program is contingent on adequate Federal funding to complement state investments in this activity. This standardized death certificate coding of occupation and industry of decedent is used for occupational health surveillance and the study of occupational mortality differentials (Harry H. Rosenberg, National Center for Health Statistics, report review comments, Hyattsville, Maryland, October 1983).

Thus, research on the connection between environmental exposures and cancer and heart and lung diseases would be enhanced by easier and more imaginative linkage of administrative records. The Congress is concerned about reducing the incidence of and mortality from chronic diseases arising from environmental exposures. Therefore, to improve the efficiency of

epidemiologic research studies, the Congress needs to improve a few of the current record systems of the Federal Government, weaken the barriers to specifically defined record linkage for research purposes by qualified and approved researchers, and provide legal confidentiality protection for research data collected in studies of disease etiology.

Current Activities in Linking and Improving Access to Federal Information

Federal Agencies are not simply waiting for Congress to enact legislation to improve their ability to link information and thus contribute to the prevention of illness associated with environmental exposures. Several activities, both feasibility tests of data linkage and legislative proposals to authorize controlled data linkages, are currently underway. These include

- The Bureau of the Census, NCHS, and NHLBI are evaluating the results of a feasibility study matching a sample of the Bureau's Current Population Survey to the NCHS National Death Index.[6]

- The IRS in cooperation with the SSA is determining the feasibility of coding occupation from a sample of the IRS Form 1040 series and then matching the coded records with a portion of the SSA one percent Continuous Work History Sample.[7] This work is supported through an interagency agreement with the National Cancer Institute.

---

6. This cooperative study by the Bureau of the Census, NCHS, and NHLBI is called the U.S. National Longitudinal Study of Mortality Differentials. The first match of CPS samples to the NDI to ascertain possible deaths in 1979, 1980, or 1981 was completed in July 1983. The probabilistic algorithm developed in the pilot study will be used to weed out false matches, and then information from death certificates on cause of death and other items will be obtained (Letter from Manning Feinleib, National Center for Health Statistics, report review comments, Hyattsville, Maryland, July 8, 1983).

7. This test has been completed and appears to have been successful. A progress report was presented at the American Statistical Association's annual meeting in August 1983 (Crabbe, Sailer, and Kilss, 1983). Matching of the IRS Statistics of Income Sample file (of over 300,000 individuals) to the NDI is also planned.

- The Veterans Administration has authorized increased access to military records, but only for individuals with a known exposure to a potential occupational hazard during military service.

- NHLBI attempted to match a sample of the 1980 Census long forms to the NCHS NDI. The effort was halted due to technical difficulties.

- A proposal is being developed by IRS to place a space on the new Form 1040EZ for recording taxpayer occupation (Frederick Scheuren, Internal Revenue Service, telephone interview, Washington, D.C., May 13, 1983).

- The Department of Health and Human Services has designed and promulgated throughout the Department a general model for a "routine use" disclosure statement for research studies, thereby facilitating transfer of information outside of departmental Agencies. This routine disclosure statement has already been incorporated into many PHS system notices (Federal Register 48, 143).

- The DHHS General Counsel has issued a legal opinion that information obtained by NIOSH researchers in the conduct of epidemiology and exposure studies can be protected from outside requests for release under the Freedom of Information Act.

- The FY86 Public Health Service Legislative Program contains a proposal to amend Section 6103 of the Internal Revenue Code to permit disclosure, for health research, of certain information in IRS and SSA files, under controlled conditions, and with approval of an IRB (DHHS 1984).

The problems of access to and linkage of specific data identifiable to an individual are well-understood and appreciated within the scientific research and research management communities. Many initiatives have been taken to relax the legal and administrative impediments and overcome the technical difficulties associated with providing linkages among associated data elements. Although these efforts began several years ago (as early as 1977 in the Privacy Protection Study Commission report), much still remains to be done.

## Quantification of the Costs of Legal Impediments

Many reviews have been written during the past decade concerning the need for relaxation of legal impediments and improvements in data records and record systems to eliminate technical and administrative barriers. These reviews imply that substantial cost savings in the conduct of epidemiologic studies could be realized if better records systems were available and national linkage of records were possible. However, no one has undertaken an effort to quantify either the costs of implementing such linkages or the savings to be achieved from efficient studies that would produce more results and presumably a healthier population.

As a part of the background data collection effort associated with the Report by the Subcommittee on Oversight and Investigations of the Committee on Interstate and Foreign Commerce entitled "Data Transfer Restrictions Impede Epidemiological Research," the General Accounting Office (GAO) conducted a brief preliminary study of the magnitude of the problem. A written report of GAO findings was never published. However, a GAO investigator's verbal report indicated that data transfer restrictions were not considered to be pervasive, restrictions were most serious for cohort studies, and the costs of current legal impediments are probably not very substantial because many Federal records are used not as primary sources but as secondary sources to verify information.

In recent conversations, GAO investigators qualified their previous verbal remarks and said that the earlier study was very brief and only an overview. However, the epidemiologists interviewed indicated the need for a thorough cost study. These epidemiologists stated that the savings could be substantial because Federal records, if readily available, would become one of their primary data sources.

Both the GAO investigators and the epidemiologists interviewed agreed that a thorough cost analysis would be a significant undertaking. To be complete, each type of study (cohort, case-control, cross-sectional) and each step within each study would have to be closely traced and costed. Because epidemiologic research studies are conducted over a relatively long time period, such a cost study would take several years. Study costs excluding legal and administrative barriers could not be arrived at directly, so a lengthy set of assumptions would have to be developed.

Another approach would be to select several representative types of epidemiologic studies currently in progress in the United States and attempt to identify similar studies in countries with existing national linkage systems, such as Canada, England, or Sweden. This type of effort should provide a very useful and valuable cost analysis of expected savings.

The other side of the cost ledger, implementation of national linkage systems, is equally difficult to measure. Not only must new data elements be included on some current forms but also coding procedures must be implemented, persons must be trained in the coding procedures, existing data processing programs must be changed, and personnel to handle the additional workloads must be identified or allocated.

In both approaches, only tangible costs are addressed. There is an array of intangible costs and benefits (savings) that cannot be so readily quantified:

- The costs of potentially reduced individual privacy
- The savings from reduced disease incidence and mortality

- The costs of depicting Government as "Big Brother" in the health environment
- The savings from increased productivity in the workplace through safer working conditions
- The costs of new technologies and software research to permit increased access and record linkage
- The savings from reduced health care system resource requirements and their conversion to other national priorities
- The costs to society of more than 20 years of lost information on the relationship between disease and occupational hazards, environmental agents, and lifestyle factors
- The costs to society of epidemiologic studies not performed because the time, expense, and risk of failure were too great for researchers who had to rely almost exclusively on decentralized state, local, and private resources for their study data
- The costs to society of more than 20 years of lost experience in conducting complex epidemiologic studies in the interests of public health using nationally linked data of scientific and personal health status information.

If Congress and the public are to be convinced that removal of legal, technical, and administrative barriers is critical to the long-term health of the nation, a cost study may be an associated necessity.

## Conclusion

As is evident from (1) discussions in this report, (2) reviews by epidemiologists during the past several years, and (3) Presidential Committee Reports, legal and administrative impediments to the efficient and effective conduct of epidemiologic research do exist and represent a significant problem for what is undoubtedly a valuable research tool.

What may be needed to alleviate the problem is a locus of responsibility in the U.S. Government with the authority and the funds to access independent data bases and to satisfy the specific informational needs of health researchers without compromising personal privacy concerns. This organizational locus could establish the necessary access to specifically identified data about a particular individual for a qualified researcher while maintaining the privacy of the individual and confidentiality of the data within acceptable boundaries. The concerns of the private citizen and the information needs of the nation need not be diametrically opposed; both can be served and satisfied within a well-structured, well-conceived framework.

An illustration is offered of a critically needed epidemiologic research project that would benefit from improved access through controlled record linkages to existing national data. This illustration is taken from the research needs identified at the March 1983 Task Force Workshop on Environmentally Related Nononcogenic Lung Disease:

> Chronic nononcogenic lung diseases, a major national health problem, develop slowly and at an uneven pace over decades. Of great importance is the role played by putative risk factors that may be associated with events in early life or with the occupational and ambient environment in adulthood. Expansion of longitudinal studies in this field is required to unravel the complex interrelationships among risk factors, [beginning with] the effects of environmental exposures on the growth and development of children's lungs. Investigation is needed of well-characterized cohorts from the neonatal period through at least mid-adult life. There are, however, clear, substantial difficulties associated with the establishment of long-term follow-ups of cohorts [related to] migration out of the region of study (Task Force, in press).

## PROPOSALS TO IMPROVE THE CONDUCT OF EPIDEMIOLOGIC RESEARCH

Based on the findings presented in this report, the following legislative actions are proposed to improve the efficiency and effectiveness of epidemiologic research.

Proposal A:

> Provide access to Federal tax records maintained by the SSA and IRS for authorized Federal health researchers to obtain necessary linkage data for epidemiologic research cohort and case-control subject identification.

The results of epidemiologic studies are important to regulators and legislators in establishing standards and mandating controls. Evidence of adverse effects of exposure to occupational agents is viewed more seriously when the effects have been observed in human populations, rather than just in laboratory experiments. In spite of the recognition of the importance of epidemiologic research, Congress has enacted legislation that restricts access to the information that may be required. The impact of legal impediments on the conduct of observational epidemiologic studies is substantial and extends to various aspects of their conduct (see pages 8 and 9).

Access to and improvement of SSA and IRS records could allow development of large specific occupation- and industry-wide cohort files. These files would serve as valuable statistical data bases for establishing national and regional disease surveillance capability. Health researchers could cost-effectively use the large cohorts for automated matching against existing data bases of disease incidence and mortality to monitor routinely for variations. If excessive disease incidence or mortality variations are

discovered, the records data bases could be used to identify specific individuals for follow-up in a traditional, detailed cohort research study directed toward specific disease etiology.

Proposal B:

> Access to the IRS Individual Master File for last known filing address and date of filing should be extended to those DHHS and other Federal health Agencies authorized to conduct epidemiologic research studies. Access should be available only after the research protocol has received Institutional Review Board approval. IRB approval requires incorporation of personal data confidentiality and informed consent provisions into the protocol formulation.

A major component of epidemiologic cohort studies and historically based case-control studies is the tracing of the study subjects to ascertain (1) their current health status in the former and (2) the variables that could be associated with their current health status in the latter. Investigators use a great variety of sources, including state motor vehicle registrations, employment records, union records, and telephone directories. The average cost per individual for the first 90 percent of the cohort is probably between $5.00 and $10.00. For the last 10 percent, the difficult-to-locate individuals, the researcher frequently must turn to a commercial tracing company, which often charges $100.00 or more per individual.

Currently, only NIOSH health researchers have access to the existing IRS Individual Master File. (Note: NIOSH researchers can obtain a taxpayer's address; but not the last filing date from which the address was obtained.) The cost for IRS address information provided to NIOSH for occupationally related epidemiologic studies is about $0.10 per name. However, the costs for researchers in other health Agencies that do not have the legal access

granted NIOSH are from 50 to 1,000 times more than they would be if confidential data needed for tracing cohort and case-control subjects were available for interagency sharing.

Proposal C:

> Enact legislation to modify the Privacy Act in accordance with the recommendations of the 1977 President's Privacy Protection Study Commission, particularly permitting interagency record data linkages under proper safeguards, to improve the effectiveness and efficiency of conducting epidemiologic research.

Many Federal organizations currently collect information on individuals that, if legally accessible to other Agencies by identifiers specific to an individual, would be valuable in the conduct of epidemiologic studies. (See listing of Agencies and their relevant data collections on pages 17 to 19.) Linkage of specifically predefined data pertinent to epidemiologic studies would improve both the efficiency and effectiveness of such studies.

Records maintained by Federal Agencies contain a wealth of information about individuals, which, if linked together, could be of great value to health researchers looking for associations between exposures and disease incidence. Individual identifying information, which is essertial for follow-up and record linkage, is necessary to carry out epidemiologic studies of risk factors, natural history, and prognosis of disease and to develop new approaches for prevention, early detection, treatment, and delivery of health services.

The concern for privacy and the problems with access to and linkage of specifically defined data on individuals are well understood and appreciated by the epidemiologic researcher and within the science management community. Many efforts have been initiated to relax legal impediments and overcome

technical difficulties associated with providing linkages among associated data elements. Although these efforts began several years ago (as early as 1977 based on recommendations in the Privacy Protection Study Commission report), much remains to be done.

Proposal D:

> The National Death Index, implemented and maintained by the NCHS, should be supported to ensure its long-term continuation and improvement.

The National Death Index, described in detail in the Task Force's Fifth Annual Report to Congress (Task Force 1982, 13-16), is an important epidemiologic research tool. This central, computerized index of death record information assists investigators in determining whether persons in their studies may have died, and if so, provides the names of the states in which those deaths are recorded and the corresponding death certificate numbers. The investigator is then able to purchase copies of the death certificates from the appropriate vital statistics office and to obtain specific information, such as cause of death. Before implementation of the National Death Index, investigators often had to search the vital statistics records of all or most states to ascertain whether a death certificate had been filed for any individual in the study group. This process is very costly and time-consuming.

The National Death Index file is compiled from magnetic tapes submitted under contractual arrangement to the NCHS by the state vital statistics offices. The file begins with deaths occurring in 1979 and is updated annually. The importance of this epidemiologic data base will continue to increase in subsequent years. Adequate funds to cover the

costs of updating the file and for further development of computerized matching techniques will be an important factor toward the effective use of this new service.

Prior to 1979, mortality information was maintained by the SSA. Although these SSA systems are still used extensively by health researchers, it should be emphasized that there are several major advantages to the National Death Index. The turnaround time is much quicker than the SSA (1 month versus 6 months), and more information (e.g., death certificate number) is given by the Index, making it easier to obtain the death certificate from the state vital statistics office.

If additional historical information from the state vital statistics offices could be included (e.g., the five years prior to 1979), the system would be much more valuable. It would also then supplement the incomplete mortality data available from the SSA, especially for the latter 1970s.

## REFERENCES CITED


Beebe, G.W. 1981. Record linkage and needed improvements in existing data resources. In Banbury Report No. 9: Quantification of occupational cancer. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

_____. 1983. Prospects for national data on occupational mortality. In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

_____. 1983a. Record linkage: methodologic and legal issues. Paper presented at the Workshop on Chronic Disease in the Workplace and the Environment, Workplace Health Fund, Washington, D.C., October 31, 1983.

_____. 1983b. Long-term follow-up is a problem. Am. J. Public Health 73(3):245-246.

Crabbe, P., Sailer, P., and Kilss, B. 1983. Occupation data from tax returns: Progress report. Presented in session sponsored by the Section on Survey Research Methods, 1983 Proceedings of the American Statistical Association Annual Meeting, 312-317. Washington, D.C. In press.

DHEW. See U.S. Department of Health, Education, and Welfare.

DHHS. See U.S. Department of Health and Human Services.

Doll, R., and Peto, R. 1981. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. J. Natl. Cancer Inst. 66(6):1191-1308.

Federal Register, Vol 48, No. 143. Monday, July 25, 1983. Pps. 33748-33749.

Gordis, L., and Gold, E. 1980. Privacy, confidentiality, and the use of medical records in research. Science 207:153-56.

Higgins, I.T.T. 1983. Cross-sectional studies. In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Lloyd, J.W. 1983. Cohort or longitudinal studies? In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Mason, T.J., McKay, F.W., Hoover, R., Blot, W.J., and Fraumeni, J.F., Jr. 1975. Atlas of cancer mortality for U.S. counties, 1950-1969. DHEW Publication No. (NIH) 75-780. Washington, D.C.: Government Printing Office.

_____. 1976. Atlas of cancer mortality among U.S. non-whites for U.S. counties, 1950-1969. DHEW Publication No. (NIH) 76-1204. Washington, D.C.: Government Printing Office.

NCI. See U.S. Department of Health and Human Services. National Institutes of Health. National Cancer Institute.

NIOSH. See U.S. Department of Health and Human Services. Centers for Disease Control/National Institute for Occupational Safety and Health.

Newill, V.A. 1983. Are society's epidemiologic needs being met? In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Office of Technology Assessment. 1982. Summary. In Cancer risk, assessing and reducing the dangers in our society. A Westview Special Study. Boulder, Colorado: Westview Press, Inc.

_____. 1982a. Factors associated with cancer. In Cancer risk, assessing and reducing the dangers in our society. A Westview Special Study. Boulder, Colorado: Westview Press, Inc.

_____. 1982b. Methods for detecting and identifying carcinogens. In Cancer risk, assessing and reducing the dangers in our society. A Westview Special Study. Boulder, Colorado: Westview Press, Inc.

OPRR. See U.S. Department of Health and Human Services. National Institutes of Health. Office of the Director. Office for Protection from Research Risks.

OTA. See Office of Technology Assessment.

Public Law 95-210. Washington, D.C.: Government Printing Office.

Public Law 96-128. Veterans' Disability Compensation and Survivors' Benefit Amendments of 1979. Washington, D.C.: Government Printing Office.

Radford, E.P. 1983. Societal significance of epidemiologic studies. In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Sartwell, P.E. 1983. Case-control studies. In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Sartwell, P.E., and Last J.M. 1980. Epidemiology. In Maxcy-Rosenau public health and preventive medicine, ed. J.M. Last. 11th ed. New York: Appleton-Century-Crofts.

Task Force on Environmental Cancer and Heart and Lung Disease. Report of the workshop on environmentally related nononcogenic lung disease. March 1983, Airlie, Virginia. Envir. Res.: In press.

_____. 1982. Environmental cancer and heart and lung disease. Fifth annual report to Congress. NTIS Accession No. PB83 155762.

_____. 1981. Summary of the workshop on exposure to environmental agents, their metabolism and mechanisms of toxicity: Research needs. NTIS Accession No. PB82 212895.

U.S. Department of Health and Human Services. 1984. Availability of tax records for health research--Legislative proposal--PHS Program for FY1986. Washington D.C.: Government Printing Office.

_____. Centers for Disease Control/National Institute for Occupational Safety and Health. 1977. The right to know: Practical problems and policy issues arising from exposures to hazardous chemicals and physical agents in the workplace. Washington D.C.: Government Printing Office.

_____. Committee to Coordinate Environmental and Related Programs (DHHS/CCERP). 1983. On occupational cancer estimation. Report of the Occupational Cancer Risk Subcommittee, DHHS/CCERP, T.M. Frazier, Chairman. November 4, 1983.

_____. National Institutes of Health. National Cancer Institute. 1981. The Surveillance, Epidemiology and End Results (SEER) incidence and mortality data 1973-77. NCI Monograph #57. DHHS Publication (NIH) 81-2330.

_____. National Institutes of Health. Office of the Director. Office for Protection from Research Risks. 1983. Draft model policy of an interagency Federal response to the recommendations of the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. (Unpublished report.)

U.S. Department of Health, Education and Welfare. Public Health Service/ National Institutes of Health, National Institute of Environmental Health Sciences. 1977. Human health and the environment--some research needs. Report of the second Task Force for Research Planning in Environmental Health Science. DHEW Publication No. NIH 77-1277. Washington, D.C.: Government Printing Office.

BIBLIOGRAPHY

Acheson, E.D. 1967. Medical record linkage. The Oxford Research Linkage Study and the Unit of Clinical Epidemiology, Nuffield Department of Clinical Medicine, University of Oxford. London: Oxford University Press.

Alvey, W., and Aziz, F. (Social Security Administration). 1979. Quality of mortality reporting in SSA linked data: Some preliminary descriptive results. Presented in session sponsored by the Section on Survey Research Methods, 1979 Proceedings of the American Statistical Association Annual Meeting.

Beebe, G.W. 1980. Record linkage systems--Canada vs. the United States. Am. J. Public Health 70(12):1246-48.

Chiazze, L., Jr. 1983. Posing the question in an epidemiologic study. In Methods and issues in occupational and environmental epidemiology, ed. L. Chiazze, Jr., F.E. Lundin, and D. Watkins. Ann Arbor, Michigan: Ann Arbor Science Publishers.

Committee on Science, Engineering, and Public Policy. National Academy of Sciences. National Academy of Engineering. Institute of Medicine. 1983. Report of the research briefing panel on human health effects of hazardous chemical exposures. Washington, D.C.: National Academy Press.

Last, J.M., ed. 1980. Maxcy-Rosenau public health and preventive medicine. 11th ed. New York: Appleton-Century-Crofts.

Office of Technology Assessment. 1982. Cancer risk, assessing and reducing the dangers in our society. A Westview Special Study. Boulder, Colorado: Westview Press, Inc.

Patterson, J.E. (National Center for Health Statistics). 1983. Evaluation of the matching effectiveness of the National Death Index. Presented in session sponsored by Social Statistics Section, 1983 Proceedings of the American Statistical Association Annual Meeting.

Privacy Act of 1974. Establishment of System of Records. Federal Register, Monday, July 25, 1983, 48(143):33748-51. Washington, D.C.: Government Printing Office.

Privacy Protection Study Commission. 1977. Personal privacy in an information society. Washington, D.C.: Government Printing Office.

Rogot, E. (NHLBI), Schwartz, S., O'Conor, K., and Olson, C. (Census Bureau, Department of Commerce). 1983. The use of probabilistic methods in matching census samples to the National Death Index. Presented in session sponsored by the Section on Survey Research Methods, 1983 Proceedings of the American Statistical Association Annual Meeting. Washington, D.C.: In press.

U.S. Congress. House Subcommittee on Oversight and Investigations of the Committee on Interstate and Foreign Commerce. 1980. Data transfer restrictions impede epidemiological research. 96th Cong., 2nd sess.

U.S. Department of Health and Human Services. 1981. Data systems of the National Center for Health Statistics. In Vital and Health Statistics, Series 1 Programs and Collection Procedures, No. 16. DHHS Publication No. (PHS) 82-1318. Washington, D.C.: Government Printing Office.

_____. 1980. Environmental health: A plan for collecting and coordinating statistical and epidemiologic data. DHHS Publication No. (PHS) 80-1248. November 4, 1983. Washington, D.C.: Government Printing Office.

_____. 1981. Environmental health: A study of the issues in locating, assessing, and treating individuals exposed to hazardous substances. DHHS Publication No. (PHS) 81-1279. Washington D.C.: Government Printing Office.

_____. National Center for Health Statistics. 1979. NCHS fact sheet.

U.S. Department of Health, Education and Welfare (DHEW). 1979. Report of the interagency task force on the health effects of ionizing radiation. Office of the Secretary, HEW, Washington, D.C.

Westin, A.F., and Baker, M.A. 1972. Databases in a free society: Computers, record-keeping and privacy. New York: Quadrangle/New York Times Book Company.

# Appendix A

## LIST OF INTERVIEWEES

Appendix A

LIST OF INTERVIEWEES

Dr. Gilbert W. Beebe, Expert Scientist
Clinical Epidemiology Branch
National Cancer Institute

Dr. Aaron E. Blair, Chief
Occupational Studies Section
National Cancer Institute

Mr. David Brown, Assistant Chief
Industrywide Studies Branch
National Institute for Occupational Safety and Health

Dr. Jerry L. Coffey, Statistical Policy Analyst
Office of Information and Regulatory Affairs
Office of Management and Budget

Mr. John P. Fanning, Legislative Coordinator
Office of Health Planning and Evaluation
Public Health Service
Department of Health and Human Services

Dr. Manning Feinleib, Director
National Center for Health Statistics

Dr. Robert S. Gordon, Jr.
Special Assistant to the Director
National Institutes of Health

Mr. Thomas Jurkiewicz, GAO Evaluator
General Accounting Office

Dr. Thomas Mason, Chief
Population Studies Section
National Cancer Institute

Mr. Hugh O'Neill, DHHS Privacy Act Officer
FOI/Privacy Division, Office of the Assistant
Secretary for Public Affairs
Department of Health and Human Services

Dr. Frederick Scheuren, Director
Statistics of Income Division
Internal Revenue Service

Mr. William Sonneman, Senior Staff
Wage Document Master File System
Internal Revenue Service

# Appendix B

# GLOSSARY

Appendix B

GLOSSARY

| | |
|---|---|
| ACME | Automated Coding of Medical Entities |
| CPS | Current Population Survey conducted by Bureau of the Census |
| CWHS | Continuous Work History Sample (SSA) |
| case-control study | Individuals with the disease or condition under study (cases) are assembled and their past or present attributes are compared with an equal or greater number of individuals without the disease or condition (controls) to identify relevant risk factors. |
| cohort | A group of persons considered free of the disease under study (at the time of selection) who are defined by some common characteristic(s). |
| DHHS | U.S. Department of Health and Human Services |
| DoD | Department of Defense |
| EPA | U.S. Environmental Protection Agency |
| epidemiology | The study of the distribution and determinants of disease in populations. |
| etiology | The science of causes or origins |
| false negative | The condition in which a negative result is obtained for a given test, but the condition being tested for is in fact positive.<br><br>Example 1 related to disease: a study subject given a tine test shows no reaction (negative result), but in fact has tuberculosis.<br><br>Example 2 in the context of this report: the NDI system does not find any matches to the study subject identifiers, but in fact the system does contain the information that the study subject has died. |
| false positive | The condition in which a positive result is obtained for a given test, but the condition being tested for is in fact not positive. |

Example 1 related to disease: a study subject given a tine test obtains a positive reaction, but in fact does not have tuberculosis.

Example 2 in the context of this report: the NDI system provides a match (indicating the study subject has died), but on close examination, it is found that he or she is the wrong subject and the real study subject is still alive or has a different death certificate number.

| | |
|---|---|
| FOI | Freedom of Information (Act) |
| GAO | General Accounting Office |
| HANES | (National) Health and Nutrition Examination Survey |
| HDFP | Hypertension Detection and Follow-up Program |
| HIS | (National) Health Interview Survey |
| IRB | Institutional Review Board |
| IRS | Internal Revenue Service |
| MRFIT | Multiple Risk Factor Intervention Trials |
| mortality | The ratio of deaths to population |
| NCHS | National Center for Health Statistics |
| NCI | National Cancer Institute |
| NDI | National Death Index |
| NHLBI | National Heart, Lung, and Blood Institute |
| NIOSH | National Institute for Occupational Safety and Health |
| NOES | National Occupational Exposure Survey |
| NOHS | National Occupational Hazard Survey |
| OGC | Office of the General Counsel |
| OMB | Office of Management and Budget |
| OSHA | Occupational Safety and Health Administration |
| PHS | Public Health Service |

| | |
|---|---|
| SEER | Surveillance, Epidemiology, and End Results |
| SOC | Standard Occupation Codes |
| SOI | Statistics of Income |
| SSA | Social Security Administration |
| SSN | Social Security Number |
| VA | Veterans Administration |